

September 2010

## Explaining the National Research Council's PhD Program Assessment Methodology: A Guide for the University of Washington

William Zumeta  
Professor

Daniel J. Evans School of Public Affairs and College of Education

### Executive Summary

This document seeks to assist UW faculty and administrators in understanding the methodology of the National Research Council's forthcoming *Assessment of Research Doctorate Programs in the United States* as well as to identify some of its limitations and implications for this institution. The paper is based on the NRC's, *A Guide to the Methodology of the National Research Council's Assessment of Research Doctorate Programs in the United States*, released in draft form in July.

#### **Summary of the NRC methodology**

- Author of the study is a distinguished NRC committee, the Committee to Assess Research-Doctorate Programs, composed of 16 members, 15 of whom are academics, chaired by Jeremiah Ostriker, Professor of Astronomy and Provost Emeritus, Princeton University. (See Appendix A for the committee's membership.)
- The study spans more than **5,000 research doctorate programs in 62 fields across 221 U.S. universities**,<sup>1</sup> a substantial increase over earlier NRC studies in 1982 and 1995.
- It excludes small fields and programs and many professional schools (see below for details). The field taxonomy is shown in Appendix B below.)
- The committee collected an unprecedented range of data on **20 different quality-related elements in each field (19 in humanities fields and computer science** due to limitations in data), which was available to faculty raters, and are to be published on-line on September 28, 2010. The 20 data elements are shown in Appendix C of this report.
- The NRC committee emphasizes that their main point is to provide a wealth of quality-related data about each doctoral program that users can analyze in a variety of ways, including constructing their own program rankings based on their own weightings of the importance of the various quality-related elements. **The committee describes its own methodology for rating and ranking programs as strictly illustrative since there is no universally agreed way to assess quality.**
- As in the past, **the ratings are based on a survey of doctoral faculty's assessments of the quality of doctoral programs in their field** but there are important methodological differences from previous studies.

---

<sup>1</sup> Rankings of programs are provided for 59 fields. The 221 institutions include nine that participate only in joint programs with other universities.

- **Each faculty rater rated only 15 programs in his/her field** in an effort to ensure careful attention to each assessment. Statistical sampling and analytic procedures were used to ensure generalizability of the set of ratings obtained for each program. Forty to fifty faculty ratings were obtained for each program (average = 44).
- Raters assessed the quality of programs (excluding the rater's own) on **3 dimensions: faculty research activity; student support and outcomes; and diversity of the academic environment**. The latter two dimensions had not been explicitly assessed in previous studies.
- **Two summary measures were also constructed** for each rated program using elaborate statistical procedures reflecting alternative weightings of the multiple data elements and the three dimensions of faculty raters in each field. *One summary measure, "S," derived weights by surveying faculty raters as to the importance of each element in the abstract to their judgments about academic quality in a doctoral program, then applied these weights to each program's data. The second summary measure, "R," derived weights indirectly from assessments of quality of specific programs by samples of raters followed by multiple regression analysis of the ratings on the data elements to infer the implicit weighting of each.* These regression-derived weights were then applied to each program's data to produce the R-values.
- **Weights vary by field** according to faculty raters' assessments of the importance of different program elements in each field.
- In general, **a program's summary rating is much more strongly influenced by its ratings on the faculty research activity elements** than by its ratings on the elements within the student support and outcomes or diversity of the academic environment elements. This reflects the faculty raters' weightings of the importance of the elements.
- **The published data are ranges of rankings for each program on each of the three dimensions plus the two summary measures, rather than raw ratings or single point rankings.** The ranges are derived by taking 500 samples of groups of faculty raters' weights (S and R weights separately) and applying them to each program's data, then arraying the programs from the highest to the lowest ranking. Thus each program has 500 rankings that fall into a distribution. *The published ranking ranges for each program show the 5<sup>th</sup> and 95<sup>th</sup> percentiles of this distribution.* The committee cautions against trying to collapse these ranges of rankings into "average" rankings although the media and some users may well do this.
- The published ranges of rankings on the three dimensions (faculty research activity, student support and outcomes, and diversity of the academic environment) use the survey-based ("S") weights to derive the departmental ratings and rankings.
- Together with the quantitative data provided about programs on the multiple elements and dimensions, **users are able to analyze the elements related to rated performance** on each of the dimensions in each field. They can, for example, array programs in a particular field by rank on specific data elements (e.g., publications per faculty; citations per publication) that are weighted heavily by faculty raters in that field and explore differences in programs' standing on different elements and dimensions.

### **Limitations of the methodology**

- *Lack of timeliness-* The objective data are for **2005-06**; faculty surveys were done in **2006-07**. Faculty may have departed, new investments been made, etc. since.

- *No “authoritative” rankings-* In its effort to represent all the uncertainties and subjectivity in any quality rating and ranking process, the NRC’s emphasis on raw data and the potential for user manipulation of weightings opens the possibility of a free for all of self serving representations of “who’s on top.” This is certainly not their intent but it may be the result.
- *Inherent subjectivity of the ratings-* The ratings are inherently subjective in that they are based on faculty rater judgments, albeit better informed by factual data than in the past.
- *Use of volunteer raters-* Only doctoral faculty who volunteered served as raters. Their views could be systematically different from those of faculty who did not volunteer as well as from those of other arguably qualified groups not employed as faculty (e.g., recent doctoral alumni or those employed in nonacademic settings in the field).
- *Imputation of missing data-* There were few missing programs in the surveyed fields but about 17% of included programs did not supply data on one or two data elements. Field averages were used to impute values for these programs, which could introduce biases if the programs with missing items were systematically different (e.g., performed worse than assumed) on these items.
- *More subjective and arbitrary judgments were necessary in humanities fields-* Comprehensive publication and citation data were not available in these disciplines so faculty vitae, a nonstandard source, had to be used. Also, well-vetted but essentially arbitrary judgments had to be made about weightings of types of publications and scholarly awards in these fields.
- *Limitations of data on student support and outcomes-* Considerable difficulties were reported in obtaining comparable data across institutions on these items.
- *Effects of program size-* The faculty publications measure *divided publications by the number of program faculty* while the citations measure was computed *per publication*, unlike in previous NRC assessments. This means that **large doctoral programs like many at the UW get less advantage in terms of research activity rankings from their size alone** than was the case under prior NRC methodologies. This could lead to lower rankings for such programs than might generally be expected. The NRC methodology report notes that program size, measured by PhD output over 2002-2006, was associated with the R-based rankings **but not the S-based rankings**. The three dimensional rankings were based only on the S weights.
- *The field taxonomy used was considered imperfect* by some of the UW departmental chairs who were asked to assign their programs to the taxonomy, particularly in the life sciences. In these cases, chairs accepted designations that they felt may not include them in a truly comparable national group. These taxonomy assignment problems were also widely reported by our American Association of Universities Data Exchange (AAUDE) colleagues, so the controversy was not limited to UW. All UW program taxonomy assignments were verified in writing with the respective program chairs at the beginning of the study. UW’s Earth and Space Sciences and Biology programs had just combined from Geological Sciences and Geophysics and Botany and Zoology respectively, at the time of the data collection. With their chairs’ consent, both programs were reported and assigned based on their older designations.

# **Explaining the National Research Council's PhD Program Assessment Methodology: A Guide for the University of Washington**

**William Zumeta  
Professor**

**Daniel J. Evans School of Public Affairs and College of Education**

This document is designed to assist University of Washington faculty and administrators in understanding how the National Research Council's forthcoming *Assessment of Research Doctoral Programs in the United States* was researched and compiled. The NRC *Assessment* will be widely covered in the media and used by prospective students, funders and government bodies for years so it is important to understand how the program rankings it reports were developed, including their strengths and limitations. This paper seeks to explain the somewhat complex methodologies used in broadly understandable terms and will also comment on their limitations and a few possible implications for the UW. It is based on the information supplied in *A Guide to the Methodology of the National Research Council Assessment of Doctorate Programs*, released by the NRC in draft form in July 2009. The *Assessment* itself is scheduled to be released shortly.

## **Scope of the NRC's Study**

The 2009 study is the third in a series of PhD program quality assessments by the National Research Council<sup>1</sup>—earlier versions appeared in 1982 and 1995—and the basic approach, involving assessments of the quality of peer programs by faculty in each discipline, has earlier antecedents. The latest study has the broadest coverage yet, spanning more than 5,000 U.S. doctoral programs in 61 fields<sup>2</sup> across 222 universities. A very high proportion of all doctoral programs in these fields that met the NRC's criteria for inclusion were covered. Fields were included that had

---

<sup>1</sup> The National Research Council is described in *A Guide to the Methodology* as the “principal operating agency of the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities” (page iv). Much of its substantive work is carried out by committees composed of members of the academies and other leading authorities in pertinent fields. The Committee to Assess Research-Doctorate Programs that conducted the study discussed here included 16 members (15 of them academics) from a range of disciplines and institutions (see Appendix A for membership and affiliations). It was chaired by Jeremiah P. Ostriker, Professor of Astronomy and Provost Emeritus, Princeton University.

<sup>2</sup> Data were collected on 67 fields but five were emerging fields with too few programs for valid ratings and one, “Languages, Societies, and Cultures,” was judged too disparate to be rated as a single field while its “subfields,” e.g., romance languages, Russian studies, African studies, were each too small for valid ratings.

produced at least 500 PhDs in the five years prior to 2004-05 and in which programs existed at a minimum of 25 universities. Programs were included if they had produced at least 5 PhDs in the five years prior to 2005-06.

The field taxonomy is provided as Appendix B. Universities chose how to classify their programs based on this field taxonomy. The field taxonomy used was considered imperfect by some of the UW departmental chairs who were asked to assign their programs to the taxonomy, particularly in the life sciences. In these cases, chairs accepted designations which they felt may not include them in a comparable national group. Most of the data were collected during late 2006 and 2007 but the factual data apply to the 2005-06 academic year. In addition to the core element, the ratings of programs by faculty peers, a wealth of “countable” data were collected on 20 different indices (19 in the humanities fields) considered to be related to quality and these data played an important role in the overall assessment as well, as is explained below. Appendix C shows the measures, which are categorized into three *dimensions: faculty research activity; student support and outcomes; and student and faculty diversity*. Most of the countable data were collected from the participating universities, largely from the individual doctoral programs. An important exception was that, except in the humanities fields, data on faculty publications, citations and awards were collected from scientometric databases maintained by the Institute for Scientific Information (ISI), now a part of Thomson Scientific.<sup>3</sup> In the humanities, where there is no such comprehensive bibliographic source, the publication data were compiled from faculty curriculum vitae.<sup>4</sup>

### **Innovations in the 2009 Assessment**

Before describing the methodology, it is worth highlighting some innovations in this version compared to the prior NRC efforts. These changes were made in part in response to criticisms leveled at the earlier studies.

*Better data for faculty raters.* The raters had much more complete program data available to them than in the past and also had convenient links to program web information. Also, each rater was asked to assess only 15 programs in his or her discipline, rather than all listed programs, thus presumably facilitating greater attention to the qualities of each.

*Multidimensional ratings and rankings.* In the past raters typically assessed two dimensions, “scholarly quality of faculty” and “effectiveness of doctoral program.” Empirically, programs’ ratings on these two scales were very highly correlated suggesting that the former, presumably the raters’ primary concern and the one about which they had most knowledge, was driving the latter rating. This time a

---

<sup>3</sup> The citation count covered the years 2000-06 and included citations to papers published between 1981 and 2006.

<sup>4</sup> Books and articles published by humanities scholars between 1996 and 2006 were included (books were weighted 5 times the weight of an article).

methodology was devised to assess (and rank) each program on the three broad dimensions mentioned above as well as to provide a summary rating and ranking within the discipline.

*Accounting for uncertainty.* The NRC committee's methodology gives careful attention to the inevitable impact of several sources of uncertainty in the program ratings (variability among raters, variability in program data resulting from year to year fluctuations in measured values, and statistical variability in the estimation as explained below), and takes account of these analytically. The ultimate result is published *ranges of rankings* of programs on each of the three dimensions plus the summary measure, rather than raw ratings of the programs or point value rankings that are deemed to provide a misleading appearance of precision about where a program ranks within its field.

*The assessment report will be mostly manipulable spreadsheets.* The assessment report itself (as distinct from the methodology report) will contain little verbiage and will consist mostly of on-line spreadsheets providing not only the committee's conclusions about each doctoral program's appropriate ranking range on each dimension, but also a wealth of data on the individual measured components and how these were weighted. Thus, programs and their overseers will be able to compare performance to peers on particular characteristics, for example those judged most important in the discipline by the faculty panels. For their own use, users may alter the weights based upon what they think is more or less important, thereby shifting where individual programs rank according to their own criteria.

### **Summary Explanation of the Assessment Methodology**

The core element in the NRC's methodology continues to be the use of judgments of faculty in each discipline about the quality of doctoral programs (other than their own of course) in their field. Pools of potential faculty raters in each field came from lists of their doctoral faculty in the disciplines provided by the participating universities. Faculty willing to participate as raters were asked to so indicate on a questionnaire. As mentioned, each individual rated only 15 programs in his or her field. To decide who rated which programs a stratified randomization procedure was used such that rating panels and the programs they rated closely matched all programs and available faculty in the field by geographic region, faculty size of program, and, in the case of the rating panels, faculty rank. Nonrespondents in rating a particular program were replaced by other raters using the above criteria until each program had been rated by 40-50 raters.<sup>5</sup> Raters scored each program on their list on a 1-6 scale ranging from, "Not adequate for doctoral education" to "Distinguished." Multiple regression analysis was employed to infer the importance of each of the 20 (19 in humanities fields) measured data elements available to

---

<sup>5</sup> The average number of raters across all programs was 44. Overall, ratings were received from about 58% of those initially asked to rate any given program. Presumably, the others felt that, even with the data supplied, they knew too little about some programs to responsibly assess them.

raters in their assessments of program quality.<sup>6</sup> These “indirect weights” were computed for each discipline separately.

All faculty in each discipline were also asked to indicate how important they judged each of the 20 (or 19) measured elements, as well as each of the three broad dimensions, to be in assessing a doctoral program’s quality.<sup>7</sup> These are referred to as “direct weights” on the elements and again they are discipline specific. (See Appendix D for a list of the elements within each dimension of quality.) Because the results from the two weighting methods were found to be “similar in magnitude but not strongly correlated” (page 17) and to take account of the various sources of uncertainty, the committee used a somewhat complex statistical procedure to combine the weights to produce the final program rating and ranking ranges. Basically, a randomly sampled set of direct and regression-derived (indirect) weights, i.e., a set derived from a particular sample of faculty in each discipline, were multiplied by each program’s (standardized) values on the measured (19 or 20) variables to produce a rating for that program. Then, a second random sample of faculty-provided direct and indirect weights was applied to the program’s data to produce a second program rating. The process was repeated to generate a total of 500 sample-based ratings for each program on each of the three dimensions plus the summary rating, producing distributions of ratings on each dimension that reflect the various uncertainties mentioned above. Rankings within the discipline were then calculated for each set of program ratings, i.e., 500 sets of rankings of programs for each dimension.

What is reported in the final *Assessment* document is the interquartile range, that is, the 75<sup>th</sup> and 25<sup>th</sup> percentiles, of each program’s 500 rankings on each of the three dimensions plus the summary measure. The committee warns against reducing this data to a single ranking for each program, say, by averaging the published 75<sup>th</sup> and 25<sup>th</sup> percentile values in the program’s ranking range (although this may well be done by the media and others).<sup>8</sup> The methodology guide includes an appendix (Appendix D here) illustrating the ranges of rankings across the three dimensions and the summary measure for unnamed programs within a single discipline, economics.

Drawing upon the data across all disciplines, the committee observes that individual programs’ rankings across the three dimensions are often quite different

---

<sup>6</sup> In the regression analysis, program ratings by faculty raters were the dependent variable and were regressed on the (standardized) values of the multiple data elements within each field. The resulting inferred weights for the different elements are referred to as “indirect weights” for each of the quality elements.

<sup>7</sup> The response rate of faculty on this task was approximately 86%.

<sup>8</sup> Such a procedure would be statistically invalid for there is no basis for assuming any particular distribution of the rankings. Typical ratings of a particular program might be near the top or near the bottom of its interquartile range, not necessarily near the mean or median.

and that, reflecting both the direct and indirect weightings, the summary rankings of individual programs tend to correspond most closely to their rankings on faculty research activity. This pattern is quite consistent with past NRC rankings, which have clearly been dominated by raters' perceptions of faculty scholarly attainments. The committee points out that some users may wish to emphasize programs' standing on the student support and outcomes or diversity dimensions, however, and the current study permits this in a way that was not possible before.

### **Uses and Limitations of the NRC Study and Possible UW-Specific Implications**

As just suggested, the NRC committee has advanced possibilities in this field by developing data and assessments on important dimensions of doctoral education, such as student support and outcomes and diversity attributes of doctoral programs, that have not been available in the past. In addition, the extensive programmatic data made available for analysis in relation to peer judgments about quality and its determinants make it possible to analyze what influences these judgments on the three dimensions on a discipline specific basis. This is a substantial advance over previous, fairly arbitrary quality assessments and can help programs determine how they can "improve" in terms of these peer assessments.

Despite all the resources and efforts of brilliant scholars that have gone into it, however, the latest NRC assessment still has significant limitations. Among these must be listed the following.

*Timeliness*- The NRC committee has traded off timeliness for analytic precision and comprehensiveness in interpreting the data from faculty judgments about quality using 2005-06 data. With the report appearing in the 2009-10 academic year, four years after the data on which it is based, there is ample room for programs' circumstances to have changed considerably as a result of faculty retirements, new hires, investments and withdrawals of resources, new leadership, and the like. Thus, academic policy and student judgments based upon the assessment may be inaccurate by virtue simply of being out of date.

*Reliance on subjective judgments by academic peers* - Clearly, educational quality is a subjective concept. Quality is in the eye of a substantial consensus of ostensibly qualified beholders, albeit with extensive data in hand to inform their judgments. Whatever the biases of these judges, they are reflected in the assessments that have been so carefully studied. On this point, it is notable that only tenure-track faculty were allowed to provide quality assessments, not well qualified individuals in the respective fields employed elsewhere or, perhaps, recent program alumni.

*Possible volunteer bias*- Although standard efforts to compare volunteer faculty raters to all faculty in each discipline were undertaken and these analyses did not produce notable differences on observable characteristics, there is no guarantee that those who volunteered to be raters did not systematically differ from non-volunteers as to potentially relevant unobservables such as attitudes about what

constitutes quality in a doctoral program or, more specifically, how different attributes should be weighted in assessing quality.

*Missing data-* The committee eliminated from the assessment programs that could not or did not provide data on three or more data elements and indicates that there were very few such programs in the end. Yet, more than 17% of all doctoral programs did not provide data on one or two of the requested data elements. In these cases the missing data were replaced with imputed data based on the mean of values for all programs in the field that had provided data. This procedure could well be problematic if programs that failed to provide data would have been below average for their discipline on the missing items, a plausible hypothesis though no doubt difficult to test in the absence of data.

*Quality of data about student support and outcomes-* There were considerable reports across the country of difficulties in getting comparable data from doctoral programs about student financial support and the NRC's questions and definitions on this topic were criticized as vague. Also, efforts to collect data about student completion rates and other outcomes were not fully successful raising some questions about how valid were the data before faculty raters as they assessed the "student support and outcomes" dimension. Again, though, it should be noted that this dimension had a fairly low weight in terms of raters' summary assessment of program quality.

*Scholarly productivity data in humanities fields were somewhat subjective-* As explained, in the absence of a comprehensive database on publications and citations in the humanities fields, the assessments of scholarly productivity in them had to be based on individuals' curriculum vitae which are not standardized and could be biased in unpredictable ways. Also, weightings of books compared to articles and of scholarly awards in these fields were necessarily somewhat arbitrary. Thus, program ratings and rankings in these fields may be perceived as more subjective and controversial than those in the sciences and social sciences.

*Use of publications per faculty and citations per publication measures-* In the past, NRC quality assessments have been based upon a program's overall reputation rather than being influenced, as in the present case, by the ready availability to raters of per faculty and per publication measures. It was widely noted earlier that the aggregate approach advantaged larger programs with more faculty to be recognized, such as would be the case for many University of Washington programs. The per faculty and per publication measures used in the present study remove much of this advantage which may mean that *some UW programs may rank lower than expected even though their aggregate impact may be little different than in the past.*

*Impacts of the field taxonomy used-* The field taxonomy used was considered imperfect by some of the UW departmental chairs who were asked to assign their programs to the taxonomy, particularly in the life sciences. In these cases, chairs accepted designations which they felt may not include them in a comparable

national group. These taxonomy assignment problems were also widely reported by our American Association of Universities Data Exchange (AAUDE) colleagues, so the controversy was not limited to UW. All UW program taxonomy assignments were verified in writing with the respective program chairs at the beginning of the study.

UW's Earth and Space Sciences and Biology programs had just combined from Geological Sciences and Geophysics and Botany and Zoology respectively, at the time of the data collection. Both programs were reported, with their chairs' consent, and assigned based on their older designations.

## **Conclusion**

The limitations of any such exercise—including its fundamental subjectivity— notwithstanding, it seems that the NRC committee has made a considerable advance in the practice of doctoral program assessment in this vein over previous efforts. Much more factual program-level data is provided to both raters and users. And, both the data and the assessments cover three distinct dimensions, including student outcomes and diversity, that are widely agreed to be important but were largely ignored in previous assessment studies. The authors take pains to account for multiple sources of uncertainty and variability and try to minimize the chances that anyone will view the rankings as precise to a degree that they cannot be. User programs and their overseers can study their “performance,” as gleaned from the rankings, and that of their peers in relation to a number of measured program elements that may help them improve their perceived quality on particular dimensions. Prospective students will be in a better position than before to identify programs of high standing on characteristics of particular importance to them. The most glaring shortcoming of the study is that, for all its voluminous data collection and impressive analysis, it is based on information that is now about four years old.

## **Appendix A - Committee to Assess Research-Doctorate Programs**

**Jeremiah P. Ostriker**, Ph.D. (*Chair*), Professor of Astronomy, Department of Astrophysical Sciences, and Provost *Emeritus*, Princeton University

**Virginia S. Hinshaw**, Ph.D. (*Vice Chair*), Chancellor, University of Hawai'i at Mānoa

**Elton D. Aberle**, Ph.D., Dean and Director Emeritus, College of Agricultural and Life Sciences, University of Wisconsin

**Norman M. Bradburn**, Ph.D., Tiffany and Margaret Blake Distinguished Service Professor and Provost *Emeritus*, University of Chicago, Senior Fellow, National Opinion Research Center

**John Brauman**, Ph.D., J.G. Jackson-C.J. Jackson Professor of Chemistry, Emeritus, Department of Chemistry, Stanford University

**Jonathan R. Cole**, Ph.D., John Mitchell Mason Professor of the University, and Provost *Emeritus*, Columbia University

**Paul W. Holland**, Ph.D., Frederic M. Lord Chair in Measurement and Statistics (retired), Educational Testing Service

**Eric Kaler**, Ph.D., Provost and Senior Vice President, Stony Brook University

**Earl Lewis**, Ph.D., Provost and Executive Vice President for Academic Affairs, Emory University

**Joan Lorden**, Ph.D., Provost and Vice Chancellor for Academic Affairs, University of North Carolina at Charlotte

**Carol Lynch**, Ph.D., Dean of the Graduate School *Emerita*, University of Colorado Boulder and Director of Professional Master's Programs, Council of Graduate Schools

**Robert M. Nerem**, Ph.D., Parker H. Petit Professor and Director, Institute for Bioengineering and Bioscience, Georgia Institute of Technology

**Suzanne Ortega**, Ph.D., Provost, University of New Mexico

**Robert Spinrad**, Retired Vice President, Technology Strategy, Xerox Corporation (resigned November 2007)

**Catharine R. Stimpson**, Dean and University Professor, Graduate School of Arts and Science, New York University

**Richard Wheeler**, Vice Provost, University of Illinois at Urbana-Champaign *Staff*

**Charlotte Kuh**, Ph.D., Study Director, Board on Higher Education and Workforce, and Deputy Executive Director, Division on Policy and Global Affairs

**Peter Henderson**, Ph.D., Director, Board on Higher Education and Workforce

**James Voytuk**, Ph.D., Senior Program Officer, Board on Higher Education and Workforce

**John Sislin**, Ph.D., Program Officer, Board on Higher Education and Workforce

**Kara Murphy**, Research Associate, Board on Higher Education and Workforce

**Sabrina Hall**, Program Associate, Board on Higher Education and Workforce

**Rae Allen**, Administrative Assistant, Board on Higher Education and Workforce

## Appendix B – Field Taxonomy

### TAXONOMY OF FIELDS

The **definition of a program** is a unit that satisfies at least three out of the following four criteria:

- 1) enrolls students,
- 2) has a designated faculty,
- 3) develops a curriculum for doctoral study, and
- 4) recommends students for the award of a doctoral degree.

Do not split or combine existing doctoral programs. To be listed, a program must have produced 5 Ph.D.s during the period 2001/2 to 2005/6.

Emerging Fields. Some fields are listed in this category because they are sometimes offered through stand-alone programs or are sometimes programs that are included as part of another larger field. Institutions are requested to report programs in emerging fields regardless of whether or not they are included in another program. If it is a separate program, respondents will be asked to indicate whether it is a stand-alone program or not. Other fields are listed because they have been experiencing significant growth, although they do not yet meet the size criterion for inclusion in the study. Institutions will also be asked to list the number of doctoral faculty members associated with the program and the number of students enrolled in it. Emerging fields will not be included in the ratings process.

Note: Field names are provided so that the study may group doctoral programs that are comparable to one another. Doctoral programs should be assigned to fields. A field may contain more than one program within the same institution. If a program has research strength outside its field (e.g. astrophysics in a physics program) or is interdisciplinary in nature, there will be an opportunity to select multiple fields for a program on the program questionnaire. In the case of “umbrella” programs in the biosciences, a doctoral program should be assigned to the field in which it awards a degree, not the program which admits doctoral students.

### LIFE SCIENCES

Biochemistry, Biophysics, and Structural Biology  
Cell and Developmental Biology  
Ecology and Evolutionary Biology  
Public Health  
Genetics and Genomics  
Immunology and Infectious Disease  
Biology/Integrated Biology/Integrated Biomedical Sciences (Note: Use this field only if the degree field is not specialized.)  
Kinesiology  
Microbiology  
Neuroscience and Neurobiology  
Nursing  
Pharmacology, Toxicology and Environmental Health  
Physiology  
Animal Sciences  
Entomology  
Food Science  
Forestry and Forest Sciences  
Nutrition  
Plant Sciences

*Emerging Fields:*

Bioinformatics  
Biotechnology  
Systems Biology

## **PHYSICAL SCIENCES & MATHEMATICS**

Applied Mathematics  
Astrophysics and Astronomy  
Chemistry  
Computer Sciences  
Earth Sciences  
Mathematics  
Oceanography, Atmospheric Sciences and Meteorology  
Physics  
Statistics and Probability

## **ENGINEERING**

Aerospace Engineering  
Biomedical Engineering and Bioengineering  
Chemical Engineering  
Civil and Environmental Engineering  
Computer Engineering  
Electrical and Computer Engineering  
Engineering Science and Materials (not elsewhere classified)  
Materials Science and Engineering  
Mechanical Engineering  
Operations Research, Systems Engineering and Industrial Engineering

### *Emerging Fields:*

Computational Engineering  
Information Science  
Nanoscience and Nanotechnology  
Nuclear Engineering

## **SOCIAL AND BEHAVIORAL SCIENCES**

Agricultural and Resource Economics  
Anthropology  
Communication  
Economics  
Geography  
Linguistics  
Political Science  
Public Affairs, Public Policy and Public Administration  
Psychology  
Sociology

### *Emerging Fields:*

Criminology and Criminal Justice  
Science and Technology Studies  
Urban Studies and Planning

## **ARTS AND HUMANITIES**

American Studies  
Classics  
Comparative Literature  
English Language and Literature  
French and Francophone Language and Literature  
German Language and Literature  
Language, Societies, and Cultures  
History  
History of Art, Architecture and Archaeology  
Music (except performance)  
Philosophy  
Religion  
Spanish and Portuguese Language and Literature  
Theatre and Performance Studies

*Emerging Fields:*

Feminist, Gender, and Sexuality Studies  
Film Studies  
Race, Ethnicity and post-Colonial Studies  
Rhetoric and Composition

## Appendix C – Variables Used in the Ratings Calculation

**Publications per Allocated Faculty,\*u2001-2006 (Non-Humanities):** Data from the Institute for Scientific Information were used to construct this variable. It is the average over the seven years, 2000-2006, of the number of articles for each allocated faculty member divided by the total number of faculty allocated to the program. Data were obtained by matching faculty lists supplied by the programs to the ISI list of publications.

**Number of Published Books and Articles per Allocated Faculty (Humanities):** Data from resumes submitted by the humanities faculty were used to construct this variable. This variable is made up of two measures; the number of published books and the number of articles published during the period 1986 to 2006 that were listed on the resume. The calculated measure was the sum of five times the number of books plus the number articles for each allocated faculty member divided by the faculty allocated to the program. In computing the allocated faculty to the program, only the allocations of the faculty who submitted resumes were added to get the allocation.

**Average Citations per Publication (Non-Humanities):** Data from the Institute for Scientific Information were used to construct this variable. It is the per-year average of the number of allocated citations in the years 2000-2006 to papers published during the period 1981-2006 by program faculty divided by the allocated publications that could contribute to the citations. For example, the number of allocated citations for a faculty member in 2003 is found by taking the 2003 citations to that faculty member's publications between 1981 and 2003. These counts are summed over the entire faculty in the program and divided by the sum of the allocated publications to the program in 2003.

**Percent of Faculty with Grants:** Data from the faculty questionnaire were used to construct this variable. The faculty questionnaire asks whether a faculty member's work is currently supported by an extramural grant of contract (E1). The total of faculty who answered this question in the affirmative was divided by the total respondents in the program and the percentage was calculated.

\*rBecause many faculty members supervise dissertations in more than one program, faculty members were allocated across the programs that they were associated with so that the total, taken across all programs, equaled one.

**Percent Interdisciplinary:** Data from the program questionnaire were used for this variable. Faculty were identified as either core, new, or associated. Percent interdisciplinary is the ratio of associated to the sum of core, new, and associated faculty. Allocations were not used in the construction of this variable.

**Percent Non-Asian Minority Faculty of Core and New Faculty, 2006\*\*:** Data from the program questionnaire were used for this variable. For each program the data reported for question B7, the race/ethnicity of core and new faculty in the program,

was used to compute the ratio of non-Hispanic Blacks, Hispanic, and American Indians or Alaska Natives to that of non-Hispanic Whites, non-Hispanic Blacks, Hispanic, Asian or Pacific Islanders, and American Indians or Alaska Natives. Faculty with Race/Ethnicity Unknown were excluded from the ratio. Allocations were not used in the construction of this variable.

**Percent Female Faculty of Core and New Faculty, 2006:** Data from the program questionnaire were used for this variable. For each program the data reported for question B5, the gender of core and new faculty in the program, was used to compute the ratio of core or new female faculty to the total of core and new faculty. Allocations were not used in the construction of this variable.

**Awards per Allocated Faculty:** Data from a review of 1,393 awards and honors from various scholarly organizations were used for this variable. The awards were identified by the committee as “Highly Prestigious” or “Prestigious” with the former given a weight of 5. The award recipients were matched to the faculty in all programs, and the total awards for a faculty member in a program was the sum of the weighted awards times the faculty member’s allocation to that program. These awards were added across the faculty in a program and divided by the total allocation of the faculty in the program.

**Average GRE, 2004-2006 (Verbal Measure for the Humanities, Quantitative Measure for All Other Fields):** Data from the program questionnaire were used for this variable. For each program, question D4 reported the average GRE verbal and quantitative scores for the 2003-2004, 2004-2005, and 2005-2006 academic years and the number of individuals who reported their scores. A weighted average was used to compute the average GRE, which was calculated by multiplying the number of individuals reporting scores by the reported average GRE score for each year, adding these three quantities and dividing by the sum of the individuals reporting scores.

**Percent Students Receiving Full Support in the First Year (Fall 2005):** Data from the program questionnaire were used for this variable. For each program question E8 reported the type of support that full-time graduate students received during the fall term each year of enrollment. For this variable the data for the first year were added for all types of support and divided by the total number of students.

\*\*p “Core” faculty are those whose primary appointment is in the doctoral program. “New” faculty are those with tenure track appointments who were appointed in 2003-2006.

**Percent First-Year Students with External Funding, 2005:** Data from the program questionnaire were used for this variable. For each program question E8 reported the type of support full-time graduate students received during fall term each year of enrollment. For this variable the data for the first year were added for support by externally funded fellowships and combinations of external fellowships and other internal support and then divided by the total number of students.

**Percent Non-Asian Minority Students, 2005:** Data from the program questionnaire were used for this variable. Question C9c reported the race/ethnicity of graduate students in the program. This was used to compute the ratio of non-Hispanic Blacks, Hispanics, and American Indians or Alaska Natives to that of non-Hispanic Whites, non-Hispanic Blacks, Hispanics, Asian or Pacific Islanders, and American Indians or Alaska Natives. Data with Race/Ethnicity Unknown were excluded from the ratio.

**Percent Female Students, 2005:** Data from the program questionnaire were used for this variable. Question C9 reported the gender of graduate students in the program. This was used to compute the percentage by taking the number of female graduate students divided by the total number of graduate students.

**Percent International Students, 2006:** Data from the program questionnaire were used for this variable. Question C9b reported the citizenship of graduate students in the program. These data were used to compute the percentage of international graduate students by taking the number with temporary visas and dividing it by the number of graduate students with known citizenship status.

**Average Annual Ph.D.'s Graduated 2002-2006:** Data from the program questionnaire were used for this variable. Question C1 reported the number of doctoral degrees awarded each academic year from 2001-2002 to 2005-2006. The average of these numbers was used for this variable. If no data were provided for a particular year, the average was taken over the years for which there were data.

**Average Completions (8-Year Completion Percentage for Humanities Fields, 6 Years for Other Fields):** Data from the program questionnaire were used for this variable. Questions C16 and C17 reported for males and females separately the number of graduate students who entered in different cohorts from 1996-1997 to 2005-2006 and the number in each cohort who completed in 3 years or less, in their 4th, 5th, 6th, 7th, 8th, 9th years, and in 10 or more years. To compute the completion rate, the number of doctoral students for a given entering cohort who completed their doctorate in 3 years or less and in their 4th, 5th, 6th years were totaled and the total was divided by the entering students in that cohort. This computation was made for each cohort that entered from 1996-1997 to 1998-1999 for the humanities and 1996-1997 to 2000-2001 for the other fields. Cohorts beyond these years were not considered, since the students could complete in a year that was after the final year 2005-2006 for which data were collected. To compute the average completion rate, an average was taken over 3 cohorts for the humanities and over 5 cohorts for other fields.

**Time to Degree (for Full- and Part-Time Graduates):** Data from the program questionnaire were used for this variable. Question C2 reported the median time to degree for full-time and part-time students. That reported number was used for this variable.

**Percent Ph.D.'s with Definite Plans for an Academic Position, 2001-2005:** Data from the National Science Foundation 2005 Doctorate Records File (DRF) were used for this variable. A crosswalk was generated between the DRF Specialty Fields of Study and the fields in the study taxonomy. Data from the DRF for 5 years (2001-2005) were matched by field and institution to the programs in the research-doctorate study. The percentage was computed by taking the number of individuals who have a signed contract or are negotiating a contract for a position at an educational institution and dividing by the number of doctorates in those years. Positions included employment and postdoctoral fellowships.

**Student Work Space:** Data from the program questionnaire were used for this variable. Question D12 reported the percentage of graduate students who have work space for their exclusive use. If reported percentage was 100 percent, then a value of 1 was given to this variable. Otherwise the value was -1.

**Health Insurance:** Data from the institutional questionnaire were used for this variable. Question A1 reported whether or not the institution provided health care insurance for its graduate students. If the response to this question was yes, then a value of 1 was given to this variable. If it was no, then the value was -1.

**Student Activities:** Data from the program questionnaire were used for this variable. Question D8 listed 18 different kinds of support for doctoral students or doctoral education. This variable is a count of the number of support mechanisms proved by the program or the institution.

## Appendix D – Weights and Variables for the Dimensional Measures

### Weights Used to Calculate the Dimensional Measures

#### Research Activity

<b>Average Weights</b>	<b>Average publications per faculty*</b>	<b>Average citations/publications</b>	<b>Percent faculty with grants</b>	<b>Awards per faculty</b>
Biological Sciences	0.30	0.31	0.36	0.13
Health Sciences	0.35	0.16	0.37	0.11
Engineering	0.29	0.25	0.29	0.17
Physical Sciences	0.28	0.26	0.29	0.17
Agricultural Sciences	0.34	0.18	0.34	0.13
Social Sciences	0.37	0.26	0.22	0.16
Humanities	0.53		0.15	0.32

\* for the humanities, the measure is of books per allocated faculty member.

#### Student Support and Outcomes

<b>Average Weights</b>	<b>Percent w/ full support</b>	<b>Average cohort completing in 6 years*</b>	<b>Time to degree full and part time</b>	<b>Placement of students</b>	<b>Program collects outcomes data</b>
Biological Sciences	0.26	0.26	0.14	0.17	0.17
Health Sciences	0.24	0.29	0.14	0.16	0.16
Engineering	0.34	0.20	0.10	0.18	0.18
Physical Sciences	0.29	0.23	0.12	0.18	0.18
Agricultural Sciences	0.28	0.23	0.12	0.18	0.18
Social Sciences	0.27	0.24	0.12	0.18	0.18
Humanities	0.29	0.25	0.11	0.17	0.17

\*For the humanities, the completion time is 8 years

### Diversity of the Academic Environment

<b>Average Weights</b>	<b>Percent core or new faculty underrepresented minority</b>	<b>Percent core or new faculty female</b>	<b>Percent students underrepresented minority</b>	<b>Percent students female</b>	<b>Percent students international</b>
Biological Sciences	0.14	0.23	0.30	0.25	0.09
Health Sciences	0.24	0.14	0.38	0.18	0.07
Engineering	0.13	0.16	0.26	0.27	0.18
Physical Sciences	0.10	0.20	0.20	0.29	0.21
Agricultural Sciences	0.15	0.17	0.30	0.22	0.15
Social Sciences	0.22	0.20	0.26	0.17	0.16
Humanities	0.21	0.24	0.20	0.18	0.16